ANALYSIS OF CHILI FRUITS DETECTION FROM STEREO CAMERA IMAGES USING YOLOV5

MUHAMMAD SHAHMIN NASHRI BIN SHAHRUL AZLAN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

ANALYSIS OF CHILI FRUITS DETECTION FROM STEREO CAMERA IMAGES USING YOLOV5

MUHAMMAD SHAHMIN NASHRI BIN SHAHRUL AZLAN

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering with Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

2022

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : ANALYSIS OF CHILI FRUITS DETECTION FROM STEREO CAMERA IMAGES USING YOLOV5
Sesi Pengajian : 2021/2022

Saya MUHAMMAD SHAHMIN NASHRI BIN SHAHRUL AZLAN mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

□ **SULIT*** (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

□ **TERHAD*** (Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan.

☑ **TIDAK TERHAD**

Disahkan oleh:

MUHAMMAD SHAHMIN NASHRI

(COP DAN TANDATANGAN PENYELIA)

TS. DR. MUHAMMAD NOORAZLAN SHAH BIN ZAINUDIN
Deputy Dean (Student Development)
Faculty of Electronics and Computer Engineering
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya
76100 Durian Tunggal, Melaka, Malaysia

Alamat Tetap: Quarters Ktmb
Blok K-4-2, Jalan
Lengkok
Abdullah,
Bangsar, 59000
Kuala Lumpur

Tarikh : 20 JUNE 2022

Tarikh : 20 June 2022

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this report entitled " ANALYSIS OF CHILI FRUITS DETECTION FROM STEREO CAMERA IMAGES USING YOLOV5" is the result of my own work except for quotes as cited in the references.

Signature : ....

Author : MUHAMMAD SHAHMIN NASHRI BIN SHAHRUL AZLAN ....

Date : June 2022 ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature : ....

Supervisor Name : TS. DR. MUHAMMAD NOORAZLAN SHAH BIN ZAINUDIN
Deputy Dean (Student Development)
Faculty of Electronics and Computer Engineering
Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya
76100 Durian Tunggal, Melaka, Malaysia

Date : 20 June 2022

DEDICATION

To my adored parents, lectures, family, and colleagues.

ABSTRACT

Chili is one of the fruits that has become as essential for cooking mostly for Malaysian's people. Eating chili provides an additional spicy taste in ancient times. There is evidence of archaeological discovery sites located in south-western Ecuador where they add a chili as an additional ingredient of food since 600 years ago, and it was one of the most important plant for growing areas on the American continent at the time as a chili farm. Less accurate for chili maturity and labour intensive, an automated approach for chili picking is prevalent. In many image recognition problems, 2D images is used. However, due to the lack of image information such depth, 2D images is considered impractical to be applied in real environment. Hence, this work aims to detect and recognize the chili fruits in order to estimate their maturity level and also for pursuing picking process. This work is expected to identify the shape and maturity of a chilli through the color using YoLov5. This work also is a part of our intention to develop a semi-autonomous chili picking robot.

ABSTRAK

Cili adalah salah satu buah yang menjadi keperluan untuk memasak kebanyakannya untuk rakyat Malaysia. Makan cili memberikan rasa pedas tambahan pada zaman dahulu. Terdapat bukti tapak penemuan arkeologi yang terletak di barat daya Ecuador di mana mereka menambah cili sebagai bahan tambahan makanan sejak 600 tahun lalu, dan ia merupakan salah satu tumbuhan terpenting untuk kawasan penanaman di benua Amerika pada masa itu sebagai ladang cili. Kurang tepat untuk kematangan cili dan intensif buruh, pendekatan automatik untuk memetik cili adalah lazim. Dalam banyak masalah pengecaman imej, imej 2D digunakan. Walau bagaimanapun, disebabkan kekurangan maklumat imej kedalaman sedemikian, imej 2D dianggap tidak praktikal untuk digunakan dalam persekitaran sebenar. Oleh itu, kerja ini bertujuan untuk mengesan dan mengenali buah cili bagi menganggar tahap kematangannya dan juga untuk mengikuti proses memetik. Karya ini diharapkan dapat mengenal pasti bentuk dan kematangan cili melalui warna menggunakan YoLov5. Kerja ini juga merupakan sebahagian daripada hasrat kami untuk membangunkan robot pemetik cili separa autonomi.

ACKNOWLEDGEMENTS

First and foremost, Alhamdulillah and thanks to Allah, the Almighty, for giving me to complete this Projek Sarjana Muda (PSM) successfully. Without the guidance of our Almighty, this project would not be completed on time. This thankful is also gratitude to FKEKK and internal PJP Grant from Universiti Teknikal Malaysia Melaka (UTeM) for the financial support. In this short period, there are a lot of obstacles and challenges that we must overcome. Throughout this project, i would like to express our utmost gratitude toward my project supervisor, Dr. Muhammad Noorazlan Shah Zainudin and Dr. Wira hidayat Bin Mohd Saad, for providing me the title and all the other information regarding the title proposed for my Projek Sarjana Muda (PSM) and all their germinal ideas and guidance throughout this project completion. Without my supervisor, i would not fully grasp the concept of this Projek Sarjana Muda (PSM), which could not have hindered our progress. Besides, i would like to thank our family and colleagues that always give a helping hand whenever necessary either directly or indirectly. Finally, I would like express my gratitude to everyone who has helped and supported me throughout Projek Sarjana Muda (PSM).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

CNN	:	Convolution Neural Network
RGB	:	Red, green and blue
YOLO	:	You Only Look Once
mAP	:	mean Average Precision
RNN	:	Recurrent Neural Networks
ANN	:	Artificial Neural Network
AI	:	Artificial Intelligence

CHAPTER 1

INTRODUCTION

1.1 Project Background

Depth camera gives the object information such as shape, localization, classification, and distance in real world by identifying the intensity of image to shows the distance of the object captured from a viewpoint[1]. The color information in depth image gives the information about the distance of object from viewpoint. In most digital cameras, an output images are a produced in 2D grid of pixels where the information, x and y axes. Initially, every pixel of images has a value associated with it that being called RGB which is red, green, and blue. The value of attribute produces from each pixel is from 0 to 255 to represent the color code, for example, the black has the point of (0,0,0) and pure bright red would be (255,0,0)[2]. A depth camera on the other hand, has an additional pixel value which have a different numerical value associated with them. An additional information presents the distance of the object from the camera or also called as depth information. Some depth cameras have both RGB and a depth system (D), which provide a pixel with all four values, or RGBD.

There are a several methods for calculating a depth, depending on the chosen an optimal operating condition. The conditions for calculation the depth is depending on user preferences such: How far that user need to see? What sort of accuracy that user need? Can it operate with multiple object? Can it operate for the outdoors? For instances, stereo depth camera uses infrared light onto a scene to improve the accuracy of the images[3]. Stereo depth camera has two sensor which is left imager and right imager that spaced a small distance apart, then from these two sensors, the depth camera will compare the distance from both. The two sensors used are depth and RGB sensors. These sensors work by improving correspondence between the two different data streams and to match the field of view between the depth sensors and the RGB sensor. Since the distance between two sensors is known, the depth information is obtained[4].

The integration of image recognition and object detection practices are frequently used in various industries such as agriculture, medical, security, etc. Image recognition identifies the objects or scenes contained within an image, while object detection identifies the instances and locations of those objects[5]. Image recognition can be used to automate such time-consuming tasks and the time taken to process the images more quickly and accurately than manual approach[6]. Image recognition is a critical technique used in a wide variety of applications and serves as the primary motivation for an invention of artificial intelligence such as deep learning for categorizing images according to their characteristics. This is particularly advantageous in e-commerce applications such as image retrieval and recommender systems. In the field of computer vision, object detection has undergone a rapid revolution. Due to its involvement in both object classification and object localization, it is one of the most difficult topics in the field of computer vision. In simple terms, the objective of this detection technique is to determine the location of objects within a given image, referred to as object localization, and the category to which each object belongs, referred to as object classification[7].

Most agriculture industries start to use an automatic technique for fruits to implement the recognition using deep learning. The model will train the network in a supervised manner, with images of the fruits serving as the input and labels for the fruits serving as the output. Following successful training, the Convolutional Neural Network (CNN) is one of the prominent models which able to predict the fruit's according to its label accurately. This idea is also can be used to develop a model which capable of recognizing and predicting the name of a fruit. Sometime when need to recognize thousands of fruit images in a short period of time, there are a variety of applications could be applied for fruit recognition. For chili as an example, deep learning CNN is used to recognize its types and categories[8].

The lifetime of chili fruits does not last long if the process of picking is not done properly. This chili would quickly being rotten if the picking process is too late. The maturity level of these chili can be known in 6 categories based on its color. The first categories are immature, the immature chili is in light green in color, and it takes a week for it to change the color into dark green and shiny. The second is mature in dark green color and shiny, they are also more durable than red chilies. The third category is quite mature, this category will change the color from green into red (start to change), it can be last stored for a week at room temperature. The fourth category is also quite mature, where chili in red color and exceeds the green color (changed in 50%). The green color in blackish and this chili cannot be stored for long period of time. The fifth category is still the same, which is quite mature, at this stage the red color has changed completely. The color is shiny and bright red and can last 1 to 2 days. The last category is over mature and usually will be used as a seed[9].

Stereo camera works in a same way to how human use two eyes by looking for depth perception. Our brains will calculate the difference between each eye to get the depth information. Objects that closer to eyes will appear to move significantly from eye to eye (or sensor to sensor), where an object in the far distance would appear to move very little. Stereo cameras can be used to create stereo views and three-dimensional images, as well as for range imaging. The distance between the lenses in a typical stereo camera (known as the intra-axial distance) is approximately 6.35 cm, though a longer base line (greater inter-camera distance) produces more extreme 3-dimensionality[10]. 3D images that adhere to the stereo camera theory can also be created more affordably by taking two images with the same camera but moving the camera a few inches left or right. If the image is edited in such a way that each eye sees a different image, the image appears to be three-dimensional. Although this method has issues with objects moving between views, it works well with still life.

Deep learning has established itself as a highly effective tool due to its capacity to handle large amounts of data. Hidden layer techniques have surpassed traditional techniques in popularity, particularly in pattern recognition. CNN are one of the most widely used model from deep neural networks categories[11]. As an example, CNN would recognize handwritten digits, detection of type of cancers, recognizing the face, etc. For handwritten recognition it primarily used in the postal sector to read zip codes, pin codes, and other unique identifiers. The critical point to remember about any deep learning model is that it requires a large amount of data and a significant amount of computing resources to train. CNN are a subclass of deep neural networks that are most frequently used to analyze visual imagery in deep learning. Deep learning has proven in variety of applications, including image and video recognition, image classification, image segmentation, medical image analysis, and natural language processing. CNN are specialized multilayer perceptron. Multilayer perceptron is referred to fully connected networks, in which each neuron in one layer is connected to every neuron in the following layer[12]. Due to their "complete connectivity," these networks are prone to overfitting data. The input to a CNN is a tensor of the form (number of heights X input inputs X input channels X input width). One of the many fascinating uses of convolutional neural networks is image classification. Aside from simple picture categorization, computer vision presents other exciting difficulties, with object detection being among the most intriguing. YOLO ("You Only Look Once") is an efficient method for real-time object detection. Unlike previous object detection methods, which repurposed classifiers to do detection, YOLO proposes the usage of an end-to-end neural network that simultaneously predicts bounding boxes and class probabilities. YOLO produces state-of-the-art results in object detection by using a fundamentally different approach than existing real-time object detection algorithms.

1.2 Problem Statement

Classification and localization

- In most object detection problems, the process for determining the object’s position generally referred to as the object localization task, is hard when 2D images are used. Not only for classifying those objects, but the detection of the correct position is crucial for implementing the detection process in real-environment situation

Object detection

- Object detection is not only able to accurately classify from its position and localize an object from its background, it's also needs to be incredibly fast at prediction time to meet the demands of video processing.

Multiple spatial scales and aspect ratios

- In any applications of object detection, object that is going to be detected may appear in a wide range of sizes and aspect ratios which is contributed to the difficulties of the process.

1.3 Objectives

1. To label the chili fruits images captured by using RGB stereo images
2. To train and analyze the object detection of labelled images using YoLov5 in term of detection accuracy.
3. To evaluate the performance of detection for different chili colors in terms of mean Average Precision (mAP).

1.4 Scope of work

This work only tackles the process of detection and identifying an object as chili fruits without estimating its maturity sizes. Matlab is used to calibrate the image captured from stereo camera. Intel Real sense SDK2.0 needs to be installed with Matlab developer package to get the Matlab wrapper. Matlab wrapper brings Intel Real Sense viewer function into Matlab. Matlab is used to capture the RGB and depth image. Using the makesense.ai in web browser to label the red and green chili. After that, training and validate the labeling image on google Colaboratoy by using

YOLOv5 function. Lastly, testing the data on the demo video to see whether red and green chili can be detected or not including accuracy for every chili.

1.5 Thesis Outline

This report is divided into five chapters. The first section is an introduction, in which the project summary, problem statement, objectives, and scope of work are explained. Chapter 2 includes information about the project that can be found in reference books, on the internet, in journals, or from other sources of information. Chapter 3 will discuss the robot motion limitation on the maximum and minimum angle of rotation for each degree of freedom. Chapter 4 will discuss the results and discussion in greater detail, while Chapter 5 will conclude this project and make some recommendations for future work.